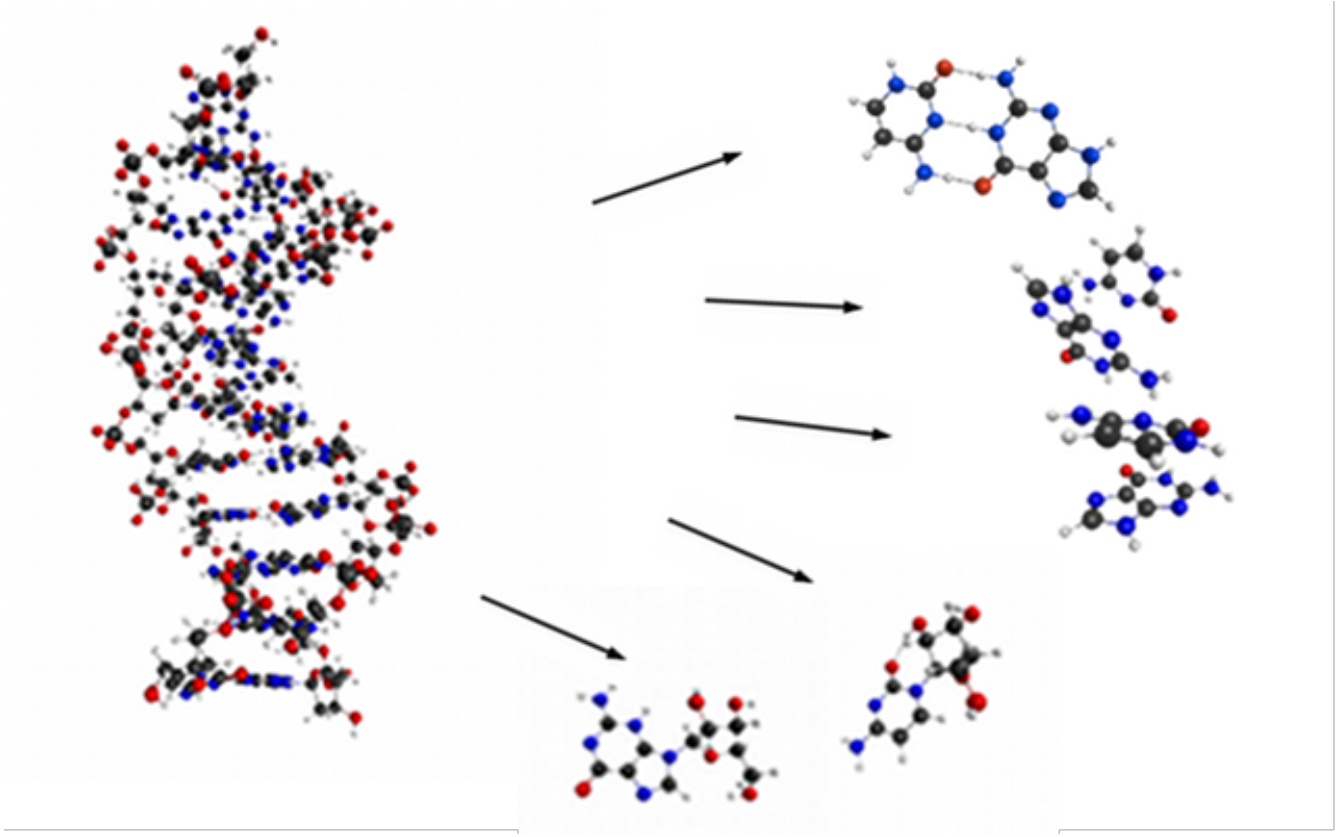
Ab initio inverstagation of the excited states of nucleobases and nucleosides

Péter G. Szalay
Laboratory of Theoretical Chemistry, Institute of Chemistry
Eötvös Loránd University, Budapest
and
Quantum Theory Project, University of Florida, Gainesville, FL

Present project

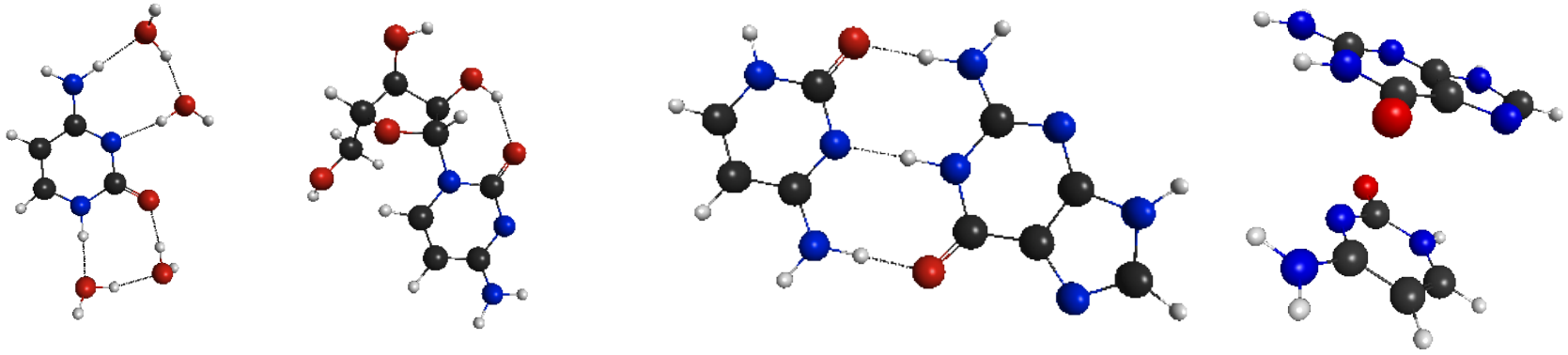
Study of spectroscopic properties of the building blocks of DNA/RNA



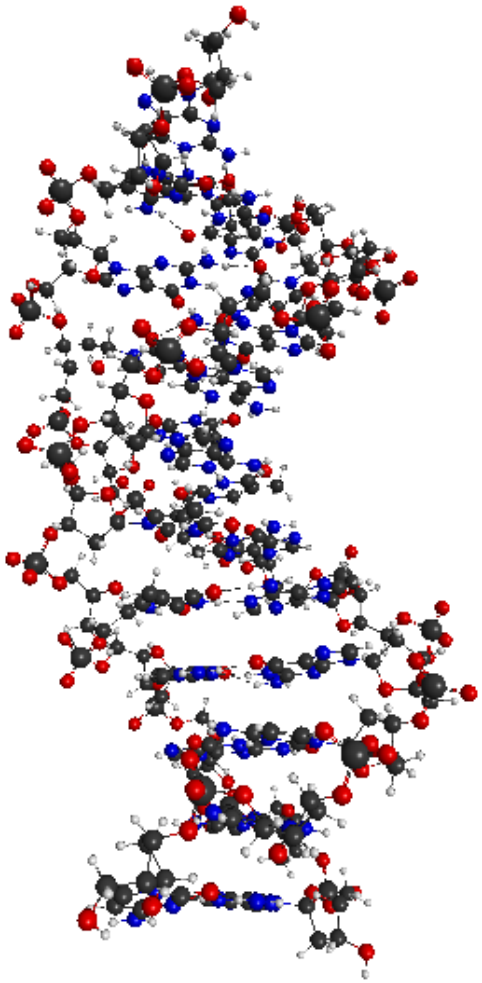
S. Lin, S. Neidle, N. Campbell, Crystal structure of the DB1880-D(CGCGAATTCGCG)$_2$ complex, Release Date: 2011-01-19, Resolution: 1.9Å, Sequence: CGCGAATTCGTCG, to be published.

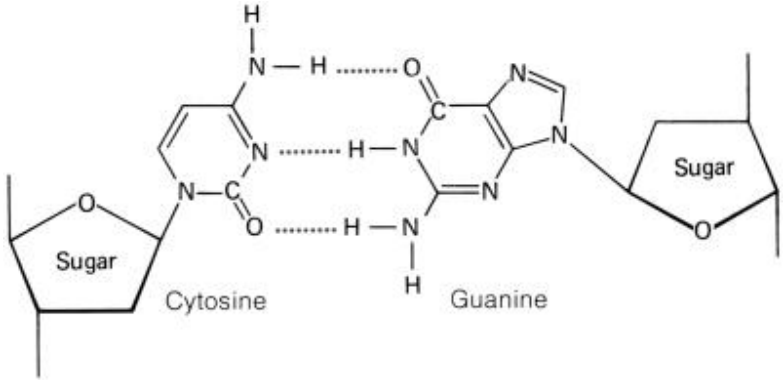
Outline of the talk

- Survey of the methods – which method should be used?
- Results:
  - How does hydration affect excited states – cytosine
  - What is the effect of sugar – cytidine and guanosine
  - Is there any change in Watson-Creek base pairs?
  - What does stacking cause?



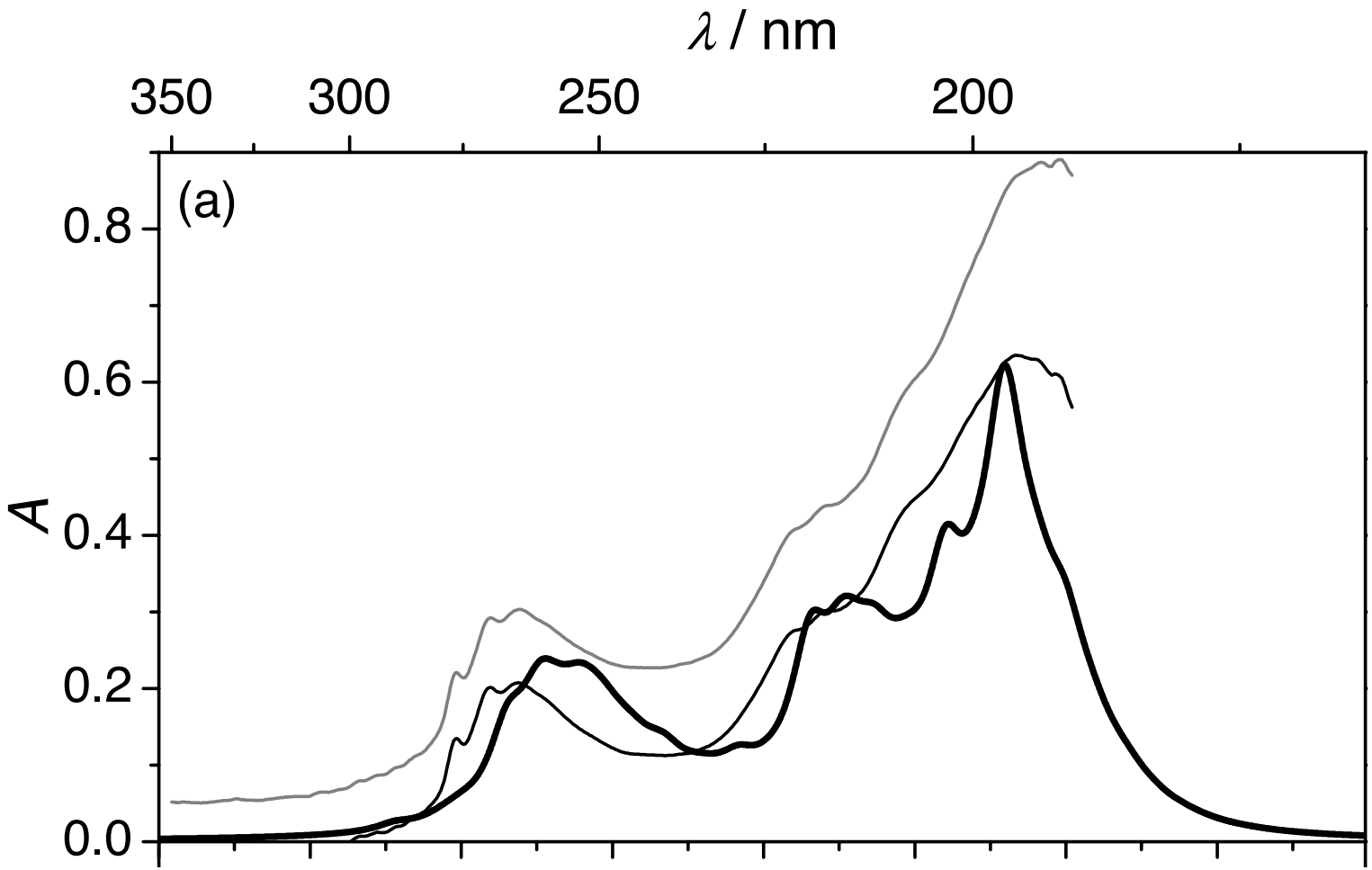
Importance of cytosine





Specific base pairs connected by hydrogen bonds

UV spectrum of cytosine
See MJ3 by Gábor Bazsó!!



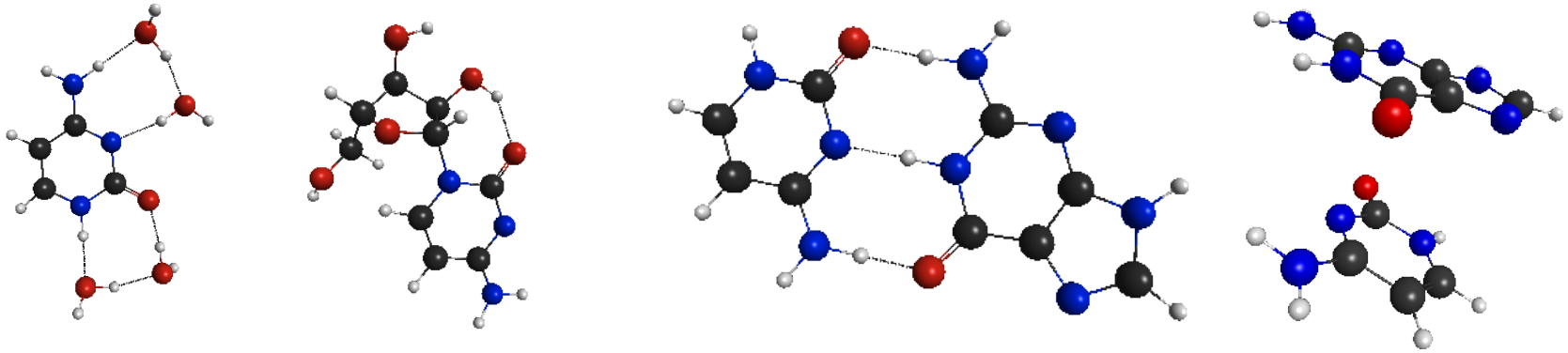
Experimental thin lines and calculated (EOM-CC) thick line

Outline of the talk

- Survey of the methods describing excited states – which method should be used?

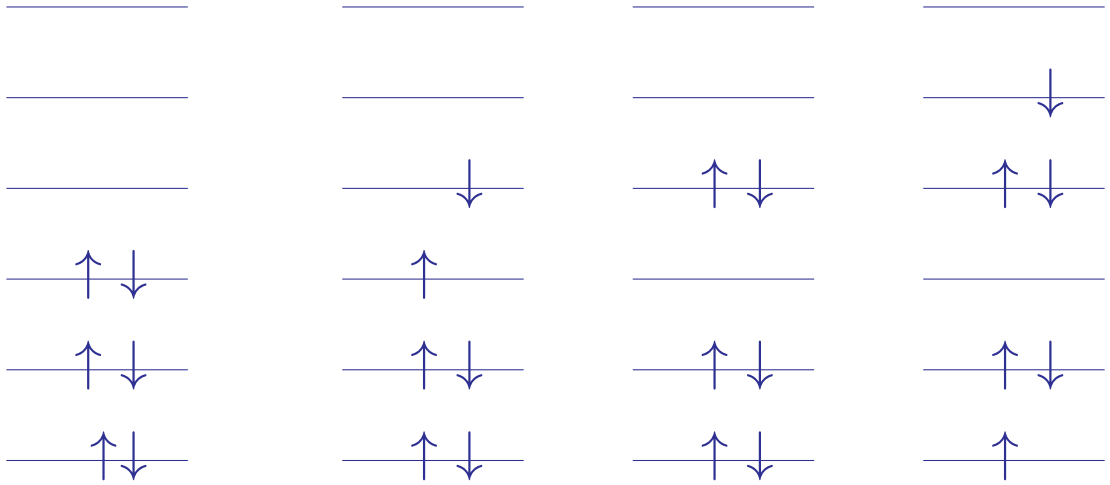
- Preliminary results:

  – How does hydration affect excited states – cytosine
  – What is the effect of sugar – cytidine and guanosine
  – Is there any change in Watson-Creek base pairs?
  – What does stacking cause?



Survey of the methods

Hierarchy of methods: depends on the excitation level included



singles doubles triples

CIS **CC2** **CCSD** **CCSD(T), CC3 ...**

Survey of the methods

Vertical excitation energy (eV) of (oxo) cytosine calculated by different methods

	EOMEE-CCSD[a]		EOMEE-CCSDT-3[b]	CC2-LR[c]		TDDFT-B3LYP[d]		CASPT2[e]	
$\pi \longrightarrow \pi^*$	4.96	0.050	4.79	4.66	0.052	4.63	0.046	4.68	0.093
$n_{N,O} \rightarrow \pi^*$	5.44	0.002	5.29	4.87	0.002	4.73	0.003	5.12	0.003
$\pi \rightarrow R$	5.73	0.004	5.52	5.53	0.005	5.11	0.001		
$\pi_N \rightarrow \pi^*$	5.89	0.148	5.65	5.61	0.138	5.44	0.074	5.54	0.352
$n \rightarrow 2\pi^*$	6.06	0.000	6.00	5.26	0.002	5.28	0.005	5.54	0.002
$\pi_N \rightarrow R$	6.23	0.007	6.05	5.83	0.000	5.65	0.001		
$\pi \rightarrow R$	6.35	0.007		5.95	0.031	5.70	0.015		
$n_{N,O} \rightarrow 2\pi^*$	6.40	0.000	5.79	6.08	0.026	6.40	0.023		
$\pi \rightarrow 2\pi^*$	6.53	0.409	6.38			6.25	0.149	6.40	0.623

[a] This work; aug-cc-pVTZ basis. [b] This work; aug-cc-pVDZ basis.

[c] Fleig, et al. JPCA, 2007; aug-cc-pVTZ basis.

[d] Shukla, Leszczynski, J. Comp. Chem. 2004; 6-311++(d,p) basis.

[e] Schreiber, Silva, Sauer Thiel, JCP, 2008. TZVP, Rydberg states are obviously missing!!!

Survey of the methods

Vertical excitation energy (eV) of (oxo) cytosine calculated by different methods

	EOMEE-CCSD[a]		EOMEE-CCSDT-3[b]	CC2-LR[c]		TDDFT-B3LYP[d]		CASPT2[e]	
$\pi \longrightarrow \pi^*$	4.96	0.050	4.79	4.66	0.052	4.63	0.046	4.68	0.093
$n_{N,O} \rightarrow \pi^*$	5.44	0.002	5.29	4.87	0.002	4.73	0.003	5.12	0.003
$\pi \rightarrow R$	5.73	0.004	5.52	5.53	0.005	5.11	0.001		
$\pi_N \rightarrow \pi^*$	5.89	0.148	5.65	5.61	0.138	5.44	0.074	5.54	0.352
$n \rightarrow 2\pi^*$	6.06	0.000	6.00	5.26	0.002	5.28	0.005	5.54	0.002
$\pi_N \rightarrow R$	6.23	0.007	6.05	5.83	0.000	5.65	0.001		
$\pi \rightarrow R$	6.35	0.007		5.95	0.031	5.70	0.015		
$n_{N,O} \rightarrow 2\pi^*$	6.40	0.000	5.79	6.08	0.026	6.40	0.023		
$\pi \rightarrow 2\pi^*$	6.53	0.409	6.38			6.25	0.149	6.40	0.623

[a] This work; aug-cc-pVTZ basis. [b] This work; aug-cc-pVDZ basis.

[c] Fleig, et al. JPCA, 2007; aug-cc-pVTZ basis.

[d] Shukla, Leszczynski, J. Comp. Chem. 2004; 6-311++(d,p) basis.

[e] Schreiber, Silva, Sauer Thiel, JCP, 2008. TZVP, Rydberg states are obviously missing!!!

Survey of the methods

Vertical excitation energy (eV) of adenine calculated by different methods

	EOMEE-CCSD[a]		+(T)[a]	CC2-LR[b]		TDDFT-B3LYP[c]		CASPT2[c]	
$\pi \rightarrow \pi^*$	5.30	0.011	5.06	5.25	–	4.98	0.205	5.20	0.146
$\pi \rightarrow 2\pi^*$	5.46	0.287	5.25	5.25	0.302	5.21	0.023	5.30	0.201
$n \rightarrow 2\pi^*$	5.54	0.006	5.39	5.12	0.007	4.88	0.013	5.21	0.001
$\pi \rightarrow R$	5.55	0.005	5.50	5.53	0.011	5.28	0.008		
$\pi \rightarrow R$	5.90	0.001	5.87	5.86	0.004	5.59	0.007		
$n \rightarrow \pi^*$	6.12	0.003	5.94	5.75	0.003	5.55	0.002	5.97	0.002
$\pi \rightarrow R$	6.38	0.001	6.38	6.08	0.030				
$n_{-1} \rightarrow 2\pi^*$	6.53	0.001	6.36	6.14	0.001	5.82	0.001		
$\pi_N \rightarrow \pi^*$	6.57	0.470	6.43			6.14	0.095	6.35	0.538
$\pi_N \rightarrow \pi^*$	6.67	0.041	6.56			6.20	0.200	6.64	0.001

[a] This work; aug-cc-pVDZ basis. [b] Fleig, et al. JPCA, 2007; aug-cc-pVTZ basis. [c] Shukla, Leszczynski, J. Comp. Chem. 2004; 6-311++(d,p) basis. [d] Schreiber, Silva, Sauer Thiel, JCP, 2008. TZVP, Rydberg states are obviously missing!!!

Survey of the methods

Vertical excitation energy (eV) of guanine calculated by different methods

	EOMEE-CCSD		+(T)	CC2-LR[a]		TDDFT-B3LYP[b]	
	aug-cc-pVDZ			aug-cc-pVTZ		6-311++(d,p)	
S_1	4.92	0.003	4.80	5.08	0.028	5.08	0.008
S_2	5.11	0.114	4.89	4.98	0.132	4.88	0.122
S_3	5.32	0.005	5.24	5.38	0.003	5.18	0.224
S_4	5.61	0.297	5.39	5.43	0.141	5.30	0.002
S_5	5.65	0.000	5.51	5.47	0.179	5.69	0.002
S_6	5.85	0.001	5.76	5.99	0.003	5.75	0.000
S_7	6.01	0.001	6.96	6.07	0.006	5.92	0.002
S_8	6.29	0.001	6.24	6.13	0.008	6.06	0.005
S_9	6.32	0.010	6.25			6.67	0.038

[a] Fleig, et al. JPCA, 2007;

[b] Shukla, Leszczynski, J. Comp. Chem. 2004;

Survey of the methods

Comparsion of the different triples methods: cytosine, aug-pVDZ basis

CCSD		CC3		CCSD(T)	CCSDT-1	CCSDT-3
4.94	0.064	4.71	0.065	4.73	5.00	4.79
5.86	0.164	5.55	0.138	5.62	5.83	5.65
6.50	0.508	6.30	0.426	6.35	6.59	6.38
6.70	0.026	6.43	0.025	6.57	6.65	6.57
6.88	0.181	6.62		6.69	6.9	6.7
7.05	0.007			6.93		
5.56	0.004	5.46	0.000	5.49	5.70	5.52
5.46	0.003	5.18	0.000	5.25	5.44	5.29
6.04	0.002	5.94	0.000	5.91	6.14	6.00
6.06	0.005	5.97	0.000	6.17	6.05	
6.19	0.007	0.00	0.000	6.08		
6.34	0.000	5.60	0.000	5.84	5.84	5.79
6.51	0.005			6.43		
6.82	0.000					

Survey of the methods

Excitation energies (ΔE in eV) and oscillator strength (f in a.u.) of cytosine and guanine dimer calculated by different methods (TZVP basis)

	ADC(2)[a]			CCSD		
	ΔE	f	assign	ΔE	f	assign
S_1	4.68	0.034	$C\pi - C\pi^*$	5.07	0.042	$C\pi - C\pi^*$
S_2	4.91	0.011	$CnG\pi - C\pi^*$	5.25	0.044	$G\pi - G\pi^*$
S_3	5.00	0.050	$G\pi - G\pi^*$	5.41	0.033	$Cn - C\pi^*$
S_4	5.21	0.128	$G\pi - C\pi^*; Cn - C\pi^*$	5.66	0.003	$Gn - G\pi^*$
S_5	5.24	0.010	$Gn - G\pi^*$	5.69	0.155	$G\pi - C\pi^*; C\pi_{-1} - C\pi$
S_6	5.25	0.010	$Cn - C\pi^*$	5.80	0.204	$G\pi - G\pi^*_{+1}$
S_7	5.46	0.231	$G\pi - G\pi^*$	5.99	0.268	$G\pi - C\pi^*; C\pi_{-1} - C\pi$
S_8	5.52	0.093	$C\pi - C\pi^*$	6.15	0.003	$Cn - C2\pi^*$
S_9	5.80	0.093	$C\pi - C\pi^*$	6.49	0.017	$G\pi - CR$
S_{10}	5.99	0.002	$Gn - G\pi^*$	6.55	0.002	$Cn_{-1} - C\pi^*$

[a] Aquino et al. (2011).

Survey of the methods

Excitation energies (ΔE in eV) and oscillator strength (f in a.u.) of adenine and thymine dimer calculated by different methods (TZVP basis)

	ADC(2)[a]			CCSD		
	ΔE	f	assign	ΔE	f	assign
S_1	4.72	0.002	$Tn - T\pi^*$	5.22	0.000	$Tn - T\pi^*$
S_2	5.06	0.018	$An - A\pi^*$	5.36	0.004	$A\pi - A\pi^*$
S_3	5.14	0.028	$A\pi - A\pi^*$	5.42	0.022	$A\pi - A\pi^*$, $T\pi - T\pi^*$
S_4	5.21	0.024	$A\pi - A\pi^*$	5.51	0.004	$An - A\pi^*$
S_5	5.32	0.347	$T\pi - T\pi^*$	5.64	0.329	$T\pi - T\pi^*$, $A\pi - A\pi^*$
S_6	5.64	0.047	$A\pi - T\pi^*$	6.14	0.002	$An - A2\pi^*$
S_7	5.77	0.004	$An - A\pi^*$	6.17	0.051	$A\pi - T\pi^*$ CT
S_8	6.10	0.002	$An - A\pi^*$	6.49	0.005	$An_{-1} - A\pi^*$
S_9	6.13	0.001	$Tn - T\pi^*$	6.60	0.086	A, $T\pi - A\pi^*$
S_{10}	6.22	0.005	$T\pi - A\pi^*$	6.67	0.008	$Tn_{-1} - T2\pi^*$
S_{11}				6.77	0.031	$T\pi - R$
S_{12}				6.77	0.031	T, $A\pi - T\pi^*$

[a] Aquino et al. (2011).

Survey of the methods

- good performance of CC2 for $\pi \rightarrow \pi^*$ states, fails for $n \rightarrow \pi^*$ and often for Rydberg states by 0.5 eV
- TDDFT works only for low lying valence states

To get a $\sim$ 0.1-0.2 eV accuracy, we have to use:

- the EOM-CCSD method
- possibly include triple excitation effects at the EOM-CCSD(T) level of theory

Survey of the methods

- good performance of CC2 for $\pi \rightarrow \pi^*$ states, fails for $n \rightarrow \pi^*$ and often for Rydberg states by 0.5 eV
- TDDFT works only for low lying valence states

To get a $\sim$ 0.1-0.2 eV accuracy, we have to use:

- the EOM-CCSD method
- possibly include triple excitation effects at the EOM-CCSD(T) level of theory

It will be very-very expensive!!!

Survey of the methods

- good performance of CC2 for $\pi \rightarrow \pi^*$ states, fails for $n \rightarrow \pi^*$ and often for Rydberg states by 0.5 eV
- TDDFT works only for low lying valence states

To get a $\sim$ 0.1-0.2 eV accuracy, we have to use:

- the EOM-CCSD method
- possibly include triple excitation effects at the EOM-CCSD(T) level of theory

It will be very-very expensive!!!

ACESIII program of the Bartlett group

ACESIII

- designed for massively parallel computers (several thousands of processors)
- http://www.qtp.ufl.edu/ACES/
- V. Lotrich, N. Flocke, M. Ponton, A. Yau, A. Perera, E. Deumens, R. J. Bartlett, Parallel Implementation of Electronic Structure Energy, Gradient and Hessian Calculations, *J. Chem. Phys*, **128**, 194104 (2008).

Calculations with ACESIII

Parameters and timings of the ACESIII calculations of this study

System	Number of				elapsed time (sec)			all job[d]
	atoms	val. els.	bf	# proc.	CCSD	EOM[b]	(T)[c]	(hour)
cytosine	13	42	229	256	180	333	820	4
adenine	15	50	275	512	325	607	1737	8.7
guanine	16	56	298	256	543	1132		
				512	363	539	1309	5.2
cytwat 1A	16	50	270	128	537	1139		4.5
cytwat 1AA	19	58	311	256	617	1373		8.5
cytwat 1W4	25	74	393	256	1931	4731		21.9
cytidine	30	94	508	512	4161	10766	36500	130
guanosine	33	108	577	512	7195	19392	96949	371
GC pair	29	98	527	512	4310	12360		
two GC pairs	58	196	1054	1024	running!			

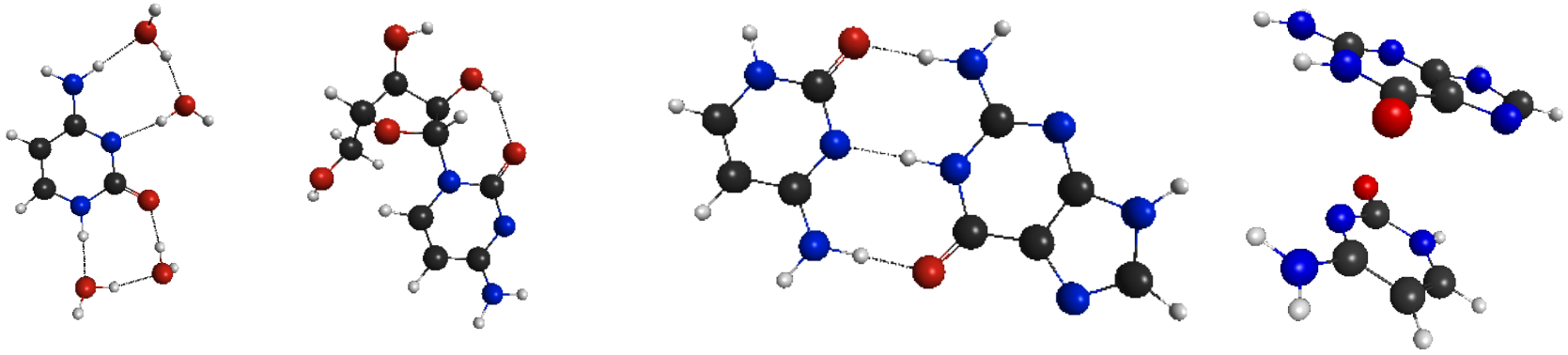
[a] Calculations have been performed with aug-cc-pVDZ basis.

[b] One root.

[c] All steps including EOM-CCSD(T), considering 12 excited states.

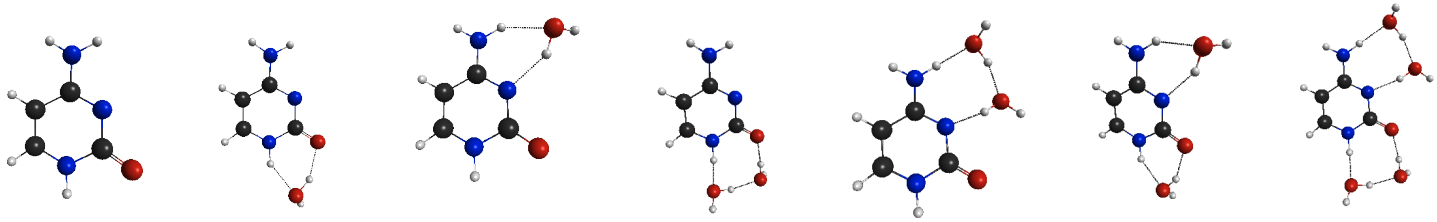
Outline of the talk

- UV spectroscopy of cytosine – how accurate can we describe it?
- Survey of the methods describing excited states – which method should be used?
- Preliminary results:
  - How does hydration affect excited states – cytosine
  - What is the effect of sugar – cytidine and guanosine
  - Is there any change in Watson-Creek base pairs?
  - What does stacking cause?



How does hydration affect excited states?

Vertical excitation energy (eV) of hydrated cytosine. "Bright states" [a)]

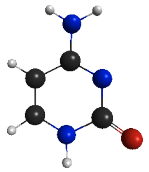
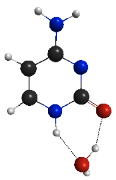
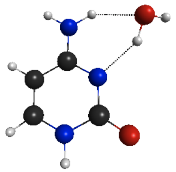
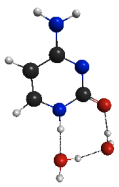
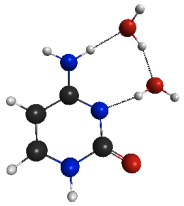
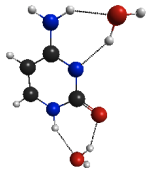
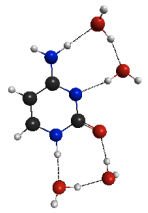


$\pi \rightarrow \pi^*$	4.94	4.85	4.88	4.88	4.93	4.93	5.00
$n_{N,O} \rightarrow \pi^*$	5.46	5.42	5.66	5.45	5.66	5.77	5.89
$\pi_N \rightarrow \pi^*$	5.86	5.83	5.69	5.85	5.64	5.72	5.71
$\pi \rightarrow 2\pi^*$	6.50	6.34	6.47	6.31	6.47	6.39	6.35
$\pi_N \rightarrow 2\pi^*$	6.88	6.81	6.77	6.80	6.63	6.72	6.63

[a)] MP2/aug-cc-pVDZ geometry, EOM-CCSD/aug-cc-pVDZ basis

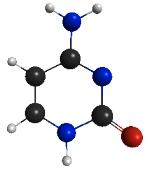
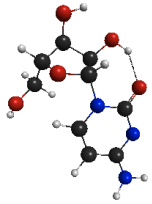
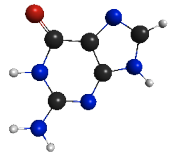
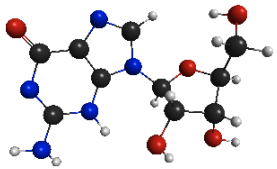
How does hydration affect excited states?

Vertical excitation energy (eV) of hydrated cytosine. "Bright" states are in red.

							
1	4.94	4.85	4.88	4.88	4.93	4.93	5.00
2	5.46	5.42	5.64	5.45	5.64	5.72	5.71
3	5.56	5.79	5.66	5.81	5.66	5.77	5.89
4	5.86	5.83	5.69	5.85	5.75	5.85	5.90
5	6.04	6.15	6.04	6.18	6.12	6.26	6.29
6	6.06	6.18	6.18	6.21	6.18	6.28	6.35
7	6.19	6.34	6.29	6.31	6.39	6.39	6.41
8	6.34	6.40	6.37	6.47	6.42	6.53	6.63
9	6.50	6.44	6.47	6.49	6.47	6.59	6.66
10	6.51	6.62	6.61	6.59	6.63	6.67	6.69
11	6.70	6.81	6.77	6.80	6.69	6.72	6.71
12	6.82	6.87	6.80	6.83	6.85	7.09	7.10

How does binded sugar affect excited states?

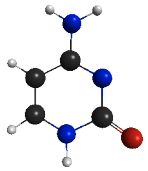
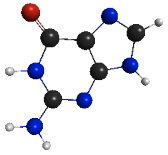
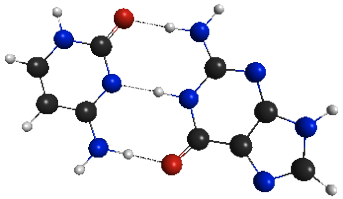
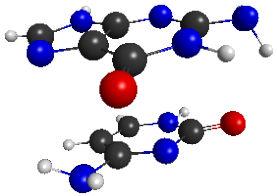
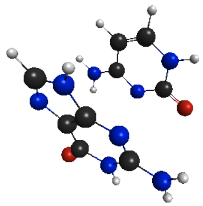
Vertical excitation energies (eV) of the "bright states" nucleobasis and nucleosides[a)]

				
$\pi \rightarrow \pi^*$	4.94	4.84	5.11	5.23
$\pi_N \rightarrow \pi^*$	5.86	5.86	–	–
$\pi \rightarrow 2\pi^*$	6.50	6.17,6.23,6.31	5.61	5.72
$\pi_N \rightarrow 2\pi^*$	6.88	6.56	–	–

[a)] MP2/aug-cc-pVDZ geometry, EOM-CCSD/aug-cc-pVDZ basis

How does stacking affect excited states?

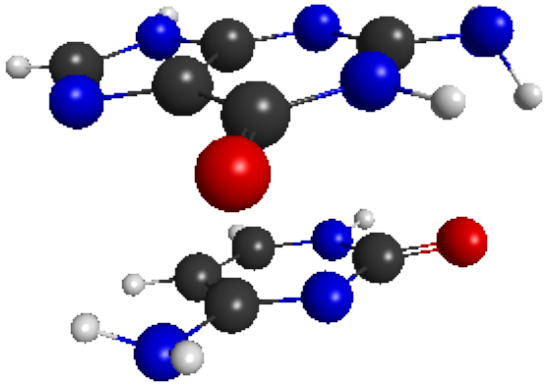
Vertical excitation energies (eV) of "bright states" of cytosine, guanine and their complexes[a)]

					
$\pi \rightarrow \pi^*$	4.94	5.11	5.07,4.89	5.00,5.15	4.81,5.05
$\pi_N \rightarrow \pi^*$	5.86	–	5.55	not seen	5.65
$\pi \rightarrow 2\pi^*$	6.50	5.61	n.c.,5.45	n.c.	n.c.
$\pi_N \rightarrow 2\pi^*$	6.88	–	n.c.	n.c.	n.c.
CT	–	–	5.68	5.54,5.63	5.53,5.80

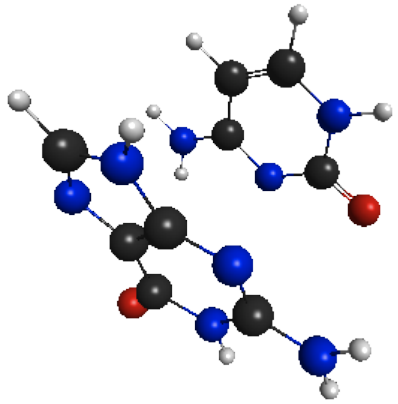
[a)] MP2/aug-cc-pVDZ geometry, EOM-CCSD/aug-cc-pVDZ basis

How does stacking affect excited states?

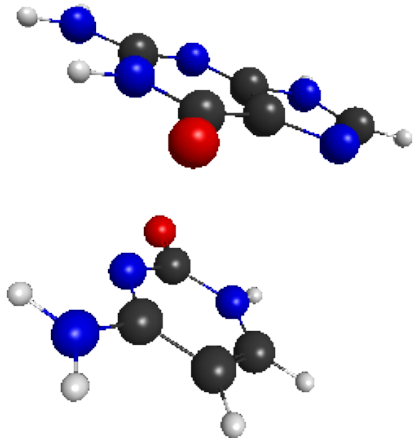
Different cytosine-guanine stacks



optimized



native 2



native 4

Conclusions

- Methodological conclusions:
  - high level ab initio methods are capable to treat systems of biological interest
  - CIS (and TD-DFT) might not be accurate enough
  - CC2 is not as accurate as it seems ($n \rightarrow \pi^*$ and Rydberg states!!)
  - EOM-CCSD and EOM-CCSD(T) provides an accuracy of 0.1-0.2 eV

- Conclusions on electronic structure:
  - surrounding waters have small effect but changes the order of excited states
  - large interaction in base pairs
  - stacking causes substantial changes: large state density, charge-transfer states

Conclusions

- Methodological conclusions:
  - high level ab initio methods are capable to treat systems of biological interest
  - CIS (and TD-DFT) might not be accurate enough
  - CC2 is not as accurate as it seems ($n \rightarrow \pi^*$ and Rydberg states!!)
  - EOM-CCSD and EOM-CCSD(T) provides an accuracy of 0.1-0.2 eV

- Conclusions on electronic structure:
  - surrounding waters have small effect but changes the order of excited states
  - large interaction in base pairs
  - stacking causes substantial changes: large state density, charge-transfer states

ACESIII is a very useful program to treat large molecules at high level!!

Acknowledgement

Co-workers:

- Prof. Géza Fogarasi (ELTE)
- Tom Watson, Dr. Ajith Perera, Dr Victor Lotrich, Prof. Rod Bartlett (UF)

Support:

- Hungarian Research Found
- HAESF (Hungarian-American Enterprise Scholarship Foundation)
- ACESIII
- Arctic Region Supercomputer Center (chugach)